

Pilot Contamination is Not a Fundamental Asymptotic Limitation in Massive MIMO

Emil Björnson*, Jakob Hoydis[†], Luca Sanguinetti^{‡§}

*Department of Electrical Engineering (ISY), Linköping University, Linköping, Sweden

[†]Nokia Bell Labs, Nozay, France.

[‡]Dipartimento di Ingegneria dell'Informazione, University of Pisa, Pisa, Italy

[§]Large Networks and System Group (LANEAS), CentraleSupélec, Université Paris-Saclay, Gif-sur-Yvette, France

Abstract—Massive MIMO provides great improvements in spectral efficiency, by coherent combining over a large antenna array and by spatial multiplexing of many users. Since its inception, the coherent interference caused by pilot contamination has been believed to be an impairment that does not vanish, even with an unlimited number of antennas. In this work, we show that this belief is incorrect and it is basically an artifact from using simplistic channel models and combining schemes. We prove that with multi-cell MMSE combining, the spectral efficiency grows without bound as the number of antennas increases, even under pilot contamination, under a condition of linear independence between the channel covariance matrices. This condition is generally satisfied, except in special cases which can be hardly found in practice.

I. INTRODUCTION

Massive MIMO (multiple-input multiple-output) is considered a key technology for the next generation of cellular networks [1]–[3], in particular, to improve the spectral efficiency (SE) and to enable spatial multiplexing of a large number of user equipments (UEs) per cell. The key difference between massive MIMO and classical multi-user MIMO is the large number of antennas, M , at each base station (BS) whose signals are processed by individual radio-frequency chains. By coherent combining, the uplink signal power of a desired UE is reinforced by a factor M , while the power of the noise and independent interference remain fixed. The same holds in the downlink. However, the pilot-based channel estimates of desired UEs are correlated with the channels to UEs that reuse the same pilots—this is known as pilot contamination. Marzetta showed in his seminal paper [1] that the interference from these UEs is also reinforced by a factor M , under the assumptions of maximum ratio combining (MRC) and independent Rayleigh fading channels. This means that pilot contamination causes the SE to have a finite limit as $M \rightarrow \infty$.

The large-antenna limit has also been studied for other combining schemes, such as the minimum mean squared error (MMSE) detector. Single-cell MMSE (S-MMSE) was considered in [4], while multi-cell MMSE (M-MMSE) was considered in [5], [6]. The difference between the M-MMSE and S-MMSE schemes is that the former makes use of channel estimates of the UEs in all cells while the latter only relies on channel estimates of the UEs in the current cell. In both cases, the SE was proved to have a finite limit as $M \rightarrow \infty$, under the

assumption of independent Rayleigh fading channels. There are special cases of spatially correlated fading that give rise to sparse channel covariance matrices. If the UEs' covariance matrices have orthogonal support, then the pilot contamination goes away and the SE grows without bound. For example, the one-ring covariance model for uniform linear arrays (ULAs), which was studied in [7], [8], gives sparse covariance matrices with orthogonal support if the channels have non-overlapping angular support. However, the ULA measurements in [9] show that such conditions are unlikely to arise in practice. This has lead us to believe that pilot contamination is a fundamental property that generally manifests a finite SE limit.

In this paper, we show that this is basically a misunderstanding, spurred by the popularity of analyzing independent Rayleigh fading channels and suboptimal combining schemes, such as MRC and S-MMSE. When using the M-MMSE combining scheme, we prove that the SE grows without bound in the presence of pilot contamination, if a simple condition of linearly independent covariance matrices is satisfied. A small amount of randomness in the covariance matrices (e.g., large-scale fading variations over the array) is sufficient to satisfy the linear independence, which makes the cases when it is not satisfied special cases rather than the general ones. We first prove this result for a simple two-user scenario in Section II and then show numerically in Section III that the result also holds for general multi-cell scenarios.

Notation: The Frobenius and spectral norms of a matrix \mathbf{X} are denoted by $\|\mathbf{X}\|_F$ and $\|\mathbf{X}\|_2$, respectively. The superscripts \top , $*$ and H denote transpose, conjugate and Hermitian transpose. We use \triangleq to denote definitions, whereas $\mathcal{N}_{\mathbb{C}}(\mathbf{x}, \mathbf{R})$ denotes the circularly symmetric complex Gaussian distribution with mean \mathbf{x} and covariance matrix \mathbf{R} . The $N \times N$ identity matrix is denoted by \mathbf{I}_N , while $\mathbf{0}_N$ is an $N \times N$ all-zero matrix. We use the shorthand $a_n \asymp b_n$ to denote $a_n - b_n \rightarrow_{n \rightarrow \infty} 0$ almost surely (a.s.) for two infinite sequences of random variables a_n, b_n .

II. PILOT CONTAMINATION IN TWO-USER SCENARIO

In this section, we prove our main result for a two-user uplink scenario, where a BS equipped with M antennas receives data from UE 1 and pilot-contaminated interference from UE 2. This setup is sufficient to demonstrate why multi-cell MMSE combining removes pilot contamination. Denote by $\mathbf{h}_k \in \mathbb{C}^M$ the channel from UE k to the BS. We consider

a Rayleigh block fading model with

$$\mathbf{h}_k \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \mathbf{R}_k), \quad k = 1, 2 \quad (1)$$

where $\mathbf{R}_k \in \mathbb{C}^{M \times M}$ is the channel covariance matrix, which is assumed to be known at the BS. The Gaussian distribution models the small-scale fading whereas the covariance matrix \mathbf{R}_k describes the macroscopic effects. The normalized trace of the covariance matrix $\beta_k = \frac{1}{M} \text{tr}(\mathbf{R}_k)$ determines the average pathloss from UE k to the BS, while the eigenstructure of \mathbf{R}_k describes the spatial channel correlation. Independent and identically distributed (i.i.d.) Rayleigh fading with $\mathbf{R}_k = \beta_k \mathbf{I}_M$ is a special case that is convenient for analysis and it is also an accurate model in isotropic fading. However, the covariance matrix has in general spatial correlation represented by non-identical diagonal elements and non-zero off-diagonal elements.

A. Channel Estimation

We assume that the BS and UEs are perfectly synchronized and operate according to a protocol wherein the uplink data transmission phase is preceded by a pilot phase for channel estimation. Both UEs use the same τ_p -length pilot sequence $\phi \in \mathbb{C}^{\tau_p}$ with elements such that $\|\phi\|^2 = \phi^H \phi = 1$. The received uplink signal $\mathbf{Y}^p \in \mathbb{C}^{N \times \tau_p}$ at the BS is given by

$$\mathbf{Y}^p = \sqrt{\rho^{\text{tr}}} \mathbf{h}_1 \phi^T + \sqrt{\rho^{\text{tr}}} \mathbf{h}_2 \phi^T + \mathbf{N}^p \quad (2)$$

where ρ^{tr} is the pilot signal-to-noise ratio (SNR) and $\mathbf{N}^p \in \mathbb{C}^{N \times \tau_p}$ is the normalized independent receiver noise with all elements distributed as $\mathcal{N}_{\mathbb{C}}(0, 1)$. The vector \mathbf{Y}^p is the observation that the BS utilizes to estimate the channels $\{\mathbf{h}_1, \mathbf{h}_2\}$. We assume that channel estimation is performed according to the MMSE estimator as given in the next lemma.

Lemma 1. *The MMSE estimator of \mathbf{h}_k for $k = 1, 2$, based on the observation \mathbf{Y}^p at the BS, is*

$$\hat{\mathbf{h}}_k = \frac{1}{\sqrt{\rho^{\text{tr}}}} \mathbf{R}_k \mathbf{Q}^{-1} \mathbf{Y}^p \phi^* \quad (3)$$

with $\mathbf{Q} = \mathbb{E}\{\mathbf{Y}^p \phi^* (\mathbf{Y}^p \phi^*)^H\} / \rho^{\text{tr}}$ being the normalized covariance matrix of the observation after correlating with the pilot sequence:

$$\mathbf{Q} = \mathbf{R}_1 + \mathbf{R}_2 + \frac{1}{\rho^{\text{tr}}} \mathbf{I}_M. \quad (4)$$

The estimate $\hat{\mathbf{h}}_k$ and the estimation error $\tilde{\mathbf{h}}_k = \mathbf{h}_k - \hat{\mathbf{h}}_k$ are independent random vectors distributed as $\tilde{\mathbf{h}}_k \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \Phi_k)$ and $\hat{\mathbf{h}}_k \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \mathbf{R}_k - \Phi_k)$ with $\Phi_k = \mathbf{R}_k \mathbf{Q}^{-1} \mathbf{R}_k$.

Proof: The proof relies on standard computations from estimation theory [10] and is omitted for space limitations. ■

The estimates $\hat{\mathbf{h}}_1$ and $\hat{\mathbf{h}}_2$ are computed in an almost identical way: the same matrix \mathbf{Q} is inverted and multiplied with the same observation $\mathbf{Y}^p \phi^* / \sqrt{\rho^{\text{tr}}}$. The only difference is that for $\hat{\mathbf{h}}_k$ there is a multiplication with the covariance matrix \mathbf{R}_k in (3), for $k = 1, 2$. The channel estimates are correlated as

$$\Upsilon_{12} = \mathbb{E}\{\hat{\mathbf{h}}_1 \hat{\mathbf{h}}_2^H\} = \mathbf{R}_1 \mathbf{Q}^{-1} \mathbf{R}_2. \quad (5)$$

If \mathbf{R}_1 is invertible, then we can also write the relation between the estimates as $\hat{\mathbf{h}}_2 = \mathbf{R}_2 \mathbf{R}_1^{-1} \hat{\mathbf{h}}_1$. In the extreme case of i.i.d. channels with $\mathbf{R}_1 = \beta_1 \mathbf{I}_M$ and $\mathbf{R}_2 = \beta_2 \mathbf{I}_M$, the two channel estimates are parallel vectors that only differ in scaling. This is an unwanted property caused by the inability of the BS to separate UEs that have transmitted the same pilot sequence over identically distributed channels. In the alternative extreme case of $\mathbf{R}_1 \mathbf{R}_2 = \mathbf{0}_M$, the two UE channels are located in completely separated subspaces, which leads to zero correlation: $\Upsilon_{12} = \mathbf{0}_M$. Consequently, it is theoretically possible to let two UEs share a pilot sequence without causing pilot contamination, if their covariance matrices satisfy the orthogonality condition $\mathbf{R}_1 \mathbf{R}_2 = \mathbf{0}_M$. In general, none of these extreme cases applies and we will investigate how to treat the partial correlation caused by pilot contamination.

We stress the fact that the MMSE estimator utilizes the (deterministic) channel statistics. In particular, the BS can only compute the MMSE estimates $\hat{\mathbf{h}}_k$ in Lemma 1 if it knows \mathbf{R}_k and also the sum of the two covariance matrices, i.e., $\mathbf{R}_1 + \mathbf{R}_2$. In practice, \mathbf{R}_k can be estimated by the sample covariance matrix, given sample realizations of \mathbf{h}_k over multiple resource blocks (e.g., different times and frequencies) where this channel is observed only in noise. Only around M samples are needed to benefit from spatial correlation in channel estimation [11].

B. Data Detection

During uplink data transmission, the received baseband signal at the BS is $\mathbf{y} \in \mathbb{C}^M$, given by

$$\mathbf{y} = \sqrt{\rho} \mathbf{h}_1 x_1 + \sqrt{\rho} \mathbf{h}_2 x_2 + \mathbf{n} \quad (6)$$

where x_k is the information-bearing signal transmitted by UE k , $\mathbf{n} \sim \mathcal{N}_{\mathbb{C}}(0, \mathbf{I}_M)$ is the independent receiver noise, and ρ is the SNR. The BS detects the signal from UE 1 by using a receive combining vector $\mathbf{v}_1 \in \mathbb{C}^M$ to obtain the scalar observation $\mathbf{v}_1^H \mathbf{y}$. Using a standard technique (see, e.g., [4]), the ergodic capacity of UE 1 is lower bounded by

$$\text{SE}_1 = \mathbb{E}\{\log_2(1 + \gamma_1)\} \quad [\text{bit/s/Hz}] \quad (7)$$

where the expectation is with respect to the channel estimates. We refer to SE_1 as a spectral efficiency. The instantaneous signal-to-interference-plus-noise ratio (SINR) γ_1 is given as

$$\begin{aligned} \gamma_1 &= \frac{|\mathbf{v}_1^H \hat{\mathbf{h}}_1|^2}{\mathbb{E}\left\{|\mathbf{v}_1^H \tilde{\mathbf{h}}_1|^2 + |\mathbf{v}_1^H \mathbf{h}_2|^2 + \frac{1}{\rho} \mathbf{v}_1^H \mathbf{v}_1 \left|\hat{\mathbf{h}}_1, \hat{\mathbf{h}}_2\right|\right\}} \\ &= \frac{|\mathbf{v}_1^H \hat{\mathbf{h}}_1|^2}{\mathbf{v}_1^H \left(\hat{\mathbf{h}}_2 \hat{\mathbf{h}}_2^H + \mathbf{Z}\right) \mathbf{v}_1} \end{aligned} \quad (8)$$

with

$$\mathbf{Z} = \sum_{k=1}^2 (\mathbf{R}_k - \Phi_k) + \frac{1}{\rho} \mathbf{I}_M. \quad (9)$$

Since γ_1 is a generalized Rayleigh quotient, it is straightforward to prove that

$$\mathbf{v}_1 = \left(\sum_{k=1}^2 \hat{\mathbf{h}}_k \hat{\mathbf{h}}_k^H + \mathbf{Z} \right)^{-1} \hat{\mathbf{h}}_1 \quad (10)$$

maximizes the SINR [5], [6]. This is called MMSE combining since (10) not only maximizes the instantaneous SINR γ_1 , but also minimizes $\mathbb{E}\{|x_1 - \mathbf{v}_1^H \mathbf{y}|^2 | \hat{\mathbf{h}}_1, \hat{\mathbf{h}}_2\}$ which is the mean squared error (MSE) in the data detection. Plugging (10) into (8) leads to

$$\gamma_1 = \hat{\mathbf{h}}_1^H \left(\hat{\mathbf{h}}_2 \hat{\mathbf{h}}_2^H + \mathbf{Z} \right)^{-1} \hat{\mathbf{h}}_1. \quad (11)$$

We will analyze how γ_1 behaves in the regime where the number of antennas, M , grows without bound, i.e., $M \rightarrow \infty$. To this end, we make the following two technical assumptions:

Assumption 1. For $k = 1, 2$,

$$\liminf_M \frac{1}{M} \text{tr}(\mathbf{R}_k) > 0 \quad (12)$$

$$\limsup_M \|\mathbf{R}_k\|_2 < \infty. \quad (13)$$

Assumption 2. Uniformly on $\lambda \in \mathbb{R}$,

$$\liminf_M \frac{1}{M} \text{tr} \left(\mathbf{Q}^{-1} (\mathbf{R}_1 - \lambda \mathbf{R}_2) \mathbf{Z}^{-1} (\mathbf{R}_1 - \lambda \mathbf{R}_2) \right) > 0. \quad (14)$$

The first assumption is a common way to model that the array gathers energy from many spatial dimensions as M grows [4], while we elaborate on the second assumption below.

The following main result is now obtained:

Theorem 1. If MMSE combining is used, then under Assumptions 1 and 2 the SINR γ_k grows a.s. unboundedly as $M \rightarrow \infty$.

Proof: The proof is given in Appendix B. ■

This theorem shows that, under certain conditions, the SE grows without bound $M \rightarrow \infty$, since a.s. $\gamma_1 \rightarrow \infty$. Observe that if the matrices \mathbf{R}_1 and \mathbf{R}_2 are linearly dependent, such that $\mathbf{R}_1 = \eta \mathbf{R}_2$, then Assumption 2 does not hold (and $\delta = 0$ in Appendix B). Under these circumstances, it is straightforward to show that $\gamma_1 \asymp \eta^2$, meaning that γ_1 converges to a finite quantity as $M \rightarrow \infty$. Next, we will elaborate on the condition that is necessary for Theorem 1.

C. Interpretation and Generality

To gain an intuitive interpretation of Assumption 2, we next provide an alternative expression that is a sufficient (but not necessary) condition for Assumption 2.

Corollary 1. Assumption 2 holds if uniformly on $\lambda \in \mathbb{R}$,

$$\liminf_M \frac{1}{M} \|\mathbf{R}_1 - \lambda \mathbf{R}_2\|_F^2 > 0. \quad (15)$$

Proof: The proof is given in Appendix C. ■

The sufficient condition in Corollary 1 requires \mathbf{R}_1 and \mathbf{R}_2 to be asymptotically linearly independent, in the sense that the difference between them grows with M . This implies that $\hat{\mathbf{h}}_1$ and $\hat{\mathbf{h}}_2$ are linearly independent. As shown in Fig. 1, it is

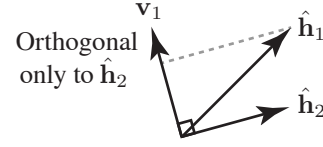


Fig. 1. If the pilot-contaminated channel estimates are linearly independent (i.e., not parallel), there exists a combining vector \mathbf{v}_1 that rejects the pilot-contaminated interference from UE 2 while $\mathbf{v}_1^H \hat{\mathbf{h}}_1 \neq 0$.

then possible to find a combining vector that is orthogonal to $\hat{\mathbf{h}}_2$, while still being partially aligned with $\hat{\mathbf{h}}_1$. This is what MMSE combining exploits to reject the pilot contamination and still achieve an array gain that grows with M .

Let us examine the condition in Corollary 1 with the help of the following two examples.

Example 1. Consider the simple scenario

$$\mathbf{R}_1 = \begin{bmatrix} 2\mathbf{I}_N & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{M-N} \end{bmatrix} \quad \mathbf{R}_2 = \mathbf{I}_M \quad (16)$$

where the covariance matrices are only different in the first N dimensions. Note that both matrices have full rank. We obtain

$$\begin{aligned} \frac{1}{M} \|\mathbf{R}_1 - \lambda \mathbf{R}_2\|_F^2 &= \frac{N(2-\lambda)^2 + (M-N)(1-\lambda)^2}{M} \\ &\geq \frac{(M-N)N}{M^2} \end{aligned} \quad (17)$$

where the inequality follows from minimizing with respect to any $\lambda \in \mathbb{R}$. Note that (17) vanishes as $M \rightarrow \infty$ if N is constant, while it has the non-zero limit $(1-\alpha)\alpha$ if $N = \alpha M$, for some α satisfying $0 < \alpha < 1$. In the latter case, the matrices $\{\mathbf{R}_1, \mathbf{R}_2\}$ satisfy (15) for all λ and thus Assumption 2 holds. Interestingly, both covariance matrices are diagonal in this example, but they are still linearly independent and the subspace where they are different has rank $\min(N, M-N)$, which is proportional to M .

Next, we study a scenario where the covariance matrices are equal except for a random perturbation. This can be interpreted as large-scale fading variations over the array.

Example 2. Consider the scenario

$$\mathbf{R}_1 = \mathbf{I}_M + \mathbf{D}_M \quad \mathbf{R}_2 = \mathbf{I}_M \quad (18)$$

where $\mathbf{D}_M = \text{diag}(d_1, \dots, d_M)$ contains i.i.d. positive random variables. This gives

$$\begin{aligned} \frac{1}{M} \|\mathbf{R}_1 - \lambda \mathbf{R}_2\|_F^2 &= \sum_{m=1}^M \frac{(d_m + 1 - \lambda)^2}{M} \\ &\asymp \mathbb{E}\{(d_m + 1 - \lambda)^2\} \geq \mathbb{E}\{(d_m - \mathbb{E}\{d_m\})^2\} \end{aligned} \quad (19)$$

by using the law of large numbers and then the fact that $\lambda - 1 = \mathbb{E}\{d_m\}$ minimizes the expression. Note that the last expression is the variance of d_m , and since every random variable has non-zero variance, we conclude that the matrices $\{\mathbf{R}_1, \mathbf{R}_2\}$ satisfy (15) and thus Assumption 2 holds.

The conclusion from Example 2 is that if we take any scenario where \mathbf{R}_1 and \mathbf{R}_2 are equal (up to a scaling factor) and then add a random perturbation to one of the matrices, then Assumption 2 holds. Hence, it is fair to say that the result of Theorem 1 holds in any non-trivial scenario.

III. PILOT CONTAMINATION IN A MULTI-CELL SCENARIO

In this section, we consider an arbitrary multi-cell scenario with L cells, each comprising a BS with M antennas and K UEs. There are $\tau_p = K$ pilot sequences and the k th UE in each cell uses the same pilot. Following the notation from [4], the received baseband signal $\mathbf{y}_j \in \mathbb{C}^M$ at BS j is

$$\mathbf{y}_j = \sum_{l=1}^L \sum_{i=1}^K \sqrt{\rho} \mathbf{h}_{jli} x_{li} + \mathbf{n}_j \quad (20)$$

where ρ is the transmit power, x_{li} is the unit-power signal from UE i in cell l , $\mathbf{h}_{jli} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \mathbf{R}_{jli})$ is the channel from this UE to BS j , $\mathbf{R}_{jli} \in \mathbb{C}^{M \times M}$ is the channel covariance matrix, and $\mathbf{n}_j \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \mathbf{I}_M)$ is the independent noise at BS j .

Using a total uplink pilot power of ρ^{tr} per UE and standard MMSE estimation techniques [4], BS j obtains the estimate

$$\hat{\mathbf{h}}_{jli} = \mathbf{R}_{jli} \mathbf{Q}_{ji}^{-1} \left(\sum_{l'=1}^L \mathbf{h}_{jl'i} + \frac{1}{\sqrt{\rho^{\text{tr}}}} \mathbf{n}_{ji} \right) \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \Phi_{jli}) \quad (21)$$

of \mathbf{h}_{jli} , where

$$\mathbf{Q}_{ji} = \sum_{l'=1}^L \mathbf{R}_{jl'i} + \frac{1}{\rho^{\text{tr}}} \mathbf{I}_M, \quad \Phi_{jli} = \mathbf{R}_{jli} \mathbf{Q}_{ji}^{-1} \mathbf{R}_{jli}. \quad (22)$$

The estimation error $\tilde{\mathbf{h}}_{jli} = \mathbf{h}_{jli} - \hat{\mathbf{h}}_{jli} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \mathbf{R}_{jli} - \Phi_{jli})$ is independent of $\hat{\mathbf{h}}_{jli}$. However, the estimates $\hat{\mathbf{h}}_{j1i}, \dots, \hat{\mathbf{h}}_{jLi}$ of the UEs with the same pilot are correlated as $\mathbb{E}\{\hat{\mathbf{h}}_{jni} \hat{\mathbf{h}}_{jmi}^H\} = \mathbf{R}_{jni} \mathbf{Q}_{ji}^{-1} \mathbf{R}_{jmi}$.

We denote by $\mathbf{v}_{jk} \in \mathbb{C}^M$ the combining vector associated with UE k in cell j . Using the same technique as in [4], the ergodic capacity of this channel is lower bounded by

$$\text{SE}_{jk} = \mathbb{E} \{ \log_2 (1 + \gamma_{jk}) \} \quad [\text{bit/s/Hz}] \quad (23)$$

with the instantaneous SINR

$$\begin{aligned} \gamma_{jk} &= \frac{|\mathbf{v}_{jk}^H \hat{\mathbf{h}}_{jjk}|^2}{\mathbb{E} \left\{ \sum_{(l,i) \neq (j,k)} |\mathbf{v}_{jk}^H \hat{\mathbf{h}}_{jli}|^2 + |\mathbf{v}_{jk}^H \tilde{\mathbf{h}}_{jjk}|^2 + \frac{1}{\rho} \mathbf{v}_{jk}^H \mathbf{v}_{jk} \left| \hat{\mathbf{h}}_{(j)} \right| \right\}} \\ &= \frac{|\mathbf{v}_{jk}^H \hat{\mathbf{h}}_{jjk}|^2}{\mathbf{v}_{jk}^H \left(\sum_{(l,i) \neq (j,k)} \hat{\mathbf{h}}_{jli} \hat{\mathbf{h}}_{jli}^H + \mathbf{Z}_j \right) \mathbf{v}_{jk}} \end{aligned} \quad (24)$$

where $\mathbb{E}\{\cdot | \hat{\mathbf{h}}_{(j)}\}$ denotes the conditional expectation given the MMSE channel estimates available at BS j and

$$\mathbf{Z}_j = \sum_{l=1}^L \sum_{i=1}^K (\mathbf{R}_{jli} - \Phi_{jli}) + \frac{1}{\rho} \mathbf{I}_M. \quad (25)$$

The following corollary finds the “optimal” receive combining vector, in the sense of maximizing SE_{jk} in (23).

Corollary 2 (see [5], [6]). *The instantaneous SINR in (24) for UE k in cell j is maximized by the combining vector*

$$\mathbf{v}_{jk} = \left(\sum_{l=1}^L \sum_{i=1}^K \hat{\mathbf{h}}_{jli} \hat{\mathbf{h}}_{jli}^H + \mathbf{Z}_j \right)^{-1} \hat{\mathbf{h}}_{jjk}. \quad (26)$$

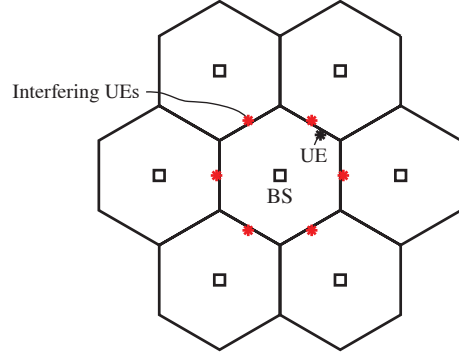


Fig. 2. Multi-cell setup with one cell-edge UE in the center cell and one cell-edge UE in each of the neighboring cells, all using the same pilot sequence.

The receive combining scheme provided by Corollary 2 is called multi-cell MMSE (M-MMSE) combining. The “multi-cell” notion is used to differentiate it from the single-cell MMSE (S-MMSE) combining scheme [4], which is widely used in the literature and is defined as

$$\mathbf{v}_{jk} = \left(\sum_{i=1}^K \hat{\mathbf{h}}_{jji} \hat{\mathbf{h}}_{jji}^H + \bar{\mathbf{Z}}_j \right)^{-1} \hat{\mathbf{h}}_{jjk}$$

with $\bar{\mathbf{Z}}_j$ being given by

$$\bar{\mathbf{Z}}_j = \sum_{i=1}^K \mathbf{R}_{jji} - \Phi_{jji} + \sum_{l=1}^L \sum_{i=1}^K \mathbf{R}_{jli} + \frac{1}{\rho} \mathbf{I}_M. \quad (27)$$

The main difference from (26) is that only channel estimates in the own cell are computed in S-MMSE, while $\hat{\mathbf{h}}_{jli} \hat{\mathbf{h}}_{jli}^H - \Phi_{jli}$ is replaced with its average (i.e., zero) for $l \neq j$. The computational complexity of S-MMSE is thus lower compared with M-MMSE, but the pilot overhead is identical since the same pilots are used to estimate both intra-cell and inter-cell channels. The S-MMSE scheme coincides with M-MMSE when there is only one isolated cell, but it is generally different and lacks the ability to suppress interference from strongly interfering UEs in other cells (e.g., located at the cell edge).

We want to analyze how γ_{jk} behaves for different combining schemes when $M \rightarrow \infty$, to show that Theorem 1 can be generalized to this multi-cell multi-user scenario. Due to the space limitation, we will only analyze this numerically.

A. Numerical Examples

To illustrate the fact that pilot contamination generally does not limit the asymptotic SE, we numerically evaluate the multi-cell scenario in Fig. 2 with $K = 1$ and $L = 7$. All UEs use the same pilot sequence and are at the cell edge near the center cell. This is a challenging setup with very high pilot contamination, and it will show our main result very clearly.

We first illustrate the eigenvalue distribution of the channel covariance matrices produced by different channel covariance models. Fig. 3 shows the ordered eigenvalues with $M = 1000$ for a covariance matrix \mathbf{R} modeled as:

1) One-ring model for a ULA with half-wavelength spacing and average pathloss β . For an angle-of-arrival (AoA) θ , the

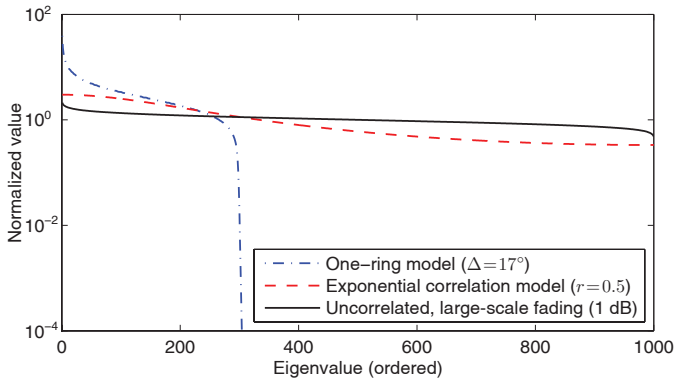


Fig. 3. Average eigenvalue distribution with three channel covariance models, whereof one gives a rank-deficient matrix and the others have full rank.

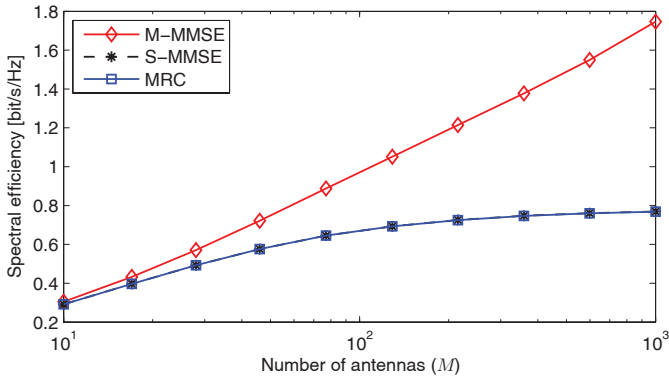


Fig. 4. SE as a function of the number of BS antennas, for covariance matrices based on the exponential correlation model in (29).

scatterers are uniformly distributed in $[\theta - \Delta, \theta + \Delta]$, which makes the (n, m) th element of \mathbf{R} become

$$[\mathbf{R}]_{m,n} = \frac{\beta}{2\Delta} \int_{-\Delta}^{\Delta} e^{\pi i(n-m) \sin(\theta+\delta)} d\delta. \quad (28)$$

2) Exponential correlation model for a ULA with correlation factor $r \in [0, 1]$ between adjacent antennas and AoA θ , which gives

$$[\mathbf{R}]_{m,n} = \beta r^{|n-m|} e^{i(n-m)\theta}. \quad (29)$$

3) Uncorrelated Rayleigh fading with independent log-normal large-scale fading over the array, which gives

$$\mathbf{R} = \beta \text{diag} \left(10^{f_1/10}, \dots, 10^{f_M/10} \right) \quad (30)$$

where $f_m \sim \mathcal{N}(0, \sigma^2)$ and σ is the standard deviation.

In Fig. 3, we show the eigenvalue spread for $\beta = 1$, $\Delta = 17^\circ$, $r = 0.5$, and $\sigma = 1$, with θ uniformly distributed in $[-\pi, +\pi]$. All three models create eigenvalue variations, but there are also substantial differences. The one-ring model provides sparse covariance matrices, where a large fraction of the eigenvalues are zero (this fraction is computed in [8]). In contrast, all eigenvalues with the other models are clearly non-zero. We consider the latter two models in the remainder to demonstrate that our main result only requires linear independence between covariance matrices, and not sparseness (which in special can give rise to orthogonal covariance supports [7]).

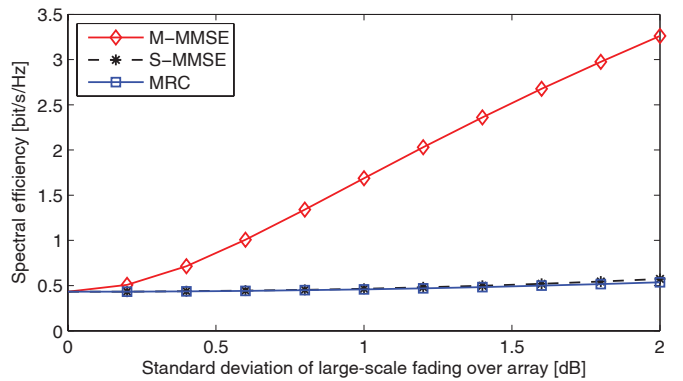


Fig. 5. SE as a function of the standard deviation of the independent large-scale fading variations, for covariance matrices modeled by (30).

The asymptotic SE behavior is considered in Fig. 4 using the exponential correlation model in (29), with M-MMSE, S-MMSE, and MRC. The average SNR observed at a BS antenna in the center cell is set equal for the pilot and data transmission: $\rho \text{tr}(\mathbf{R}_{jli})/M = \rho^{\text{tr}} \text{tr}(\mathbf{R}_{jli})/M$. It is -7.0 dB for the desired UE and -8.6 dB for each of the interfering UEs. Fig. 4 shows that S-MMSE provides slightly higher SE than MRC, but both converge to an asymptotic limit of around 0.8 bit/s/Hz as the number of antennas grows. In contrast, M-MMSE provides an SE that clearly grows without bound. The instantaneous SINR grows linearly with M , in line with our main result in Theorem 1, as seen from the fact that the SE grows linearly with a logarithmic horizontal scale.

Next, we consider the uncorrelated Rayleigh fading model in (30) with independent large-scale fading variations over the array. The SE with $M = 1000$ and varying standard deviation σ is shown in Fig. 5. M-MMSE provides no benefit over S-MMSE or MRC in the special case of $\sigma = 0$, where all covariance matrices are linearly dependent (scaled identity matrices). This is a special case that has received massive attention from researchers. However, M-MMSE provides substantial gains as soon as there are some minor variations in channel gain over the array, which effectively make the covariance matrices linearly independent. This is in line with Example 2. The range of fading variations in this simulation can be compared with the measurements in [12], which show large-scale variations of around 4 dB over a massive MIMO array.

IV. CONCLUSION

Pilot contamination generally does not cause a fundamental upper limit on the SE in massive MIMO, despite all the previous results that have pointed towards this direction. There are indeed special cases where the channel covariance matrices are linearly dependent, which make the channel estimates of the desired and interfering UEs parallel such that linear receive combining cannot remove the interference. In general, the covariance matrices and the channel estimates are not linearly dependent, thus linear M-MMSE combining can extract the desired signal while rejecting the pilot contamination. There is a power loss, as compared to the contamination-free case, but the SE still grows without bound as $M \rightarrow \infty$. Importantly,

this means that MRC (also known as matched filtering) is generally not asymptotically optimal in massive MIMO.

APPENDIX A: USEFUL RESULTS

Lemma 2 (Theorem 3.4, Corollary 3.4 [13]). *Let $\mathbf{A} \in \mathbb{C}^{M \times M}$ and $\mathbf{x}, \mathbf{y} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \frac{1}{M} \mathbf{I}_M)$. Assume that \mathbf{A} has uniformly bounded spectral norm (with respect to M) and that \mathbf{x} and \mathbf{y} are mutually independent and independent of \mathbf{A} . Then,*

$$(i) \mathbf{x}^H \mathbf{A} \mathbf{x} \asymp \frac{1}{M} \text{tr}(\mathbf{A}) \quad (ii) \mathbf{x}^H \mathbf{A} \mathbf{y} \asymp 0.$$

Lemma 3 ([14]). *For any positive semi-definite $N \times N$ matrices \mathbf{A} and \mathbf{B} , it holds that*

$$\frac{1}{N} \text{tr}(\mathbf{A} \mathbf{B}) \leq \|\mathbf{A} \mathbf{B}\|_2 \leq \|\mathbf{A}\|_2 \|\mathbf{B}\|_2. \quad (31)$$

Lemma 4 ([14]). *For any positive semi-definite $N \times N$ matrices \mathbf{A} and \mathbf{B} , it holds that*

$$\text{tr}((\mathbf{I} + \mathbf{A})^{-1} \mathbf{B}) \geq \frac{1}{1 + \|\mathbf{A}\|_2} \text{tr}(\mathbf{B}). \quad (32)$$

APPENDIX B: PROOF OF THEOREM 1

Using the matrix inversion lemma [4, Lemma 2], we may rewrite γ_1 in (11) as

$$\gamma_1 = M \left(\frac{1}{M} \hat{\mathbf{h}}_1^H \mathbf{Z}^{-1} \hat{\mathbf{h}}_1 - \frac{\left| \frac{1}{M} \hat{\mathbf{h}}_1^H \mathbf{Z}^{-1} \hat{\mathbf{h}}_2 \right|^2}{\frac{1}{M} + \frac{1}{M} \hat{\mathbf{h}}_2^H \mathbf{Z}^{-1} \hat{\mathbf{h}}_2} \right) \quad (33)$$

by also multiplying and dividing each term by M . Under Assumption 1, when $M \rightarrow \infty$, using Lemma 1 and Lemma 2 (see Appendix A) we have that¹

$$\frac{1}{M} \hat{\mathbf{h}}_1^H \mathbf{Z}^{-1} \hat{\mathbf{h}}_1 \asymp \frac{1}{M} \text{tr}(\Phi_1 \mathbf{Z}^{-1}) \triangleq \beta_{11} \quad (34)$$

$$\frac{1}{M} \hat{\mathbf{h}}_2^H \mathbf{Z}^{-1} \hat{\mathbf{h}}_2 \asymp \frac{1}{M} \text{tr}(\Phi_2 \mathbf{Z}^{-1}) \triangleq \beta_{22} \quad (35)$$

$$\frac{1}{M} \hat{\mathbf{h}}_1^H \mathbf{Z}^{-1} \hat{\mathbf{h}}_2 \asymp \frac{1}{M} \text{tr}(\Upsilon_{12} \mathbf{Z}^{-1}) \triangleq \beta_{12}. \quad (36)$$

It also follows from Assumption 1 that $\liminf_M \beta_{22} > 0$, and we then obtain

$$\frac{\gamma_1}{M} \asymp \delta = \beta_{11} - \frac{\beta_{12}^2}{\beta_{22}}. \quad (37)$$

Since Assumption 2 implies that $\liminf_M \delta > 0$ (see Appendix D) we have that γ_1 grows a.s. unboundedly and, thus, the result of the theorem follows.

APPENDIX C: PROOF OF COROLLARY 1

The expression in (14) can be lower bounded as

$$\begin{aligned} & \frac{1}{M} \text{tr}(\mathbf{Q}^{-1}(\mathbf{R}_1 - \lambda \mathbf{R}_2) \mathbf{Z}^{-1}(\mathbf{R}_1 - \lambda \mathbf{R}_2)) \\ & \geq \frac{\frac{1}{M} \text{tr}((\mathbf{R}_1 - \lambda \mathbf{R}_2)(\mathbf{R}_1 - \lambda \mathbf{R}_2))}{(\rho^{\text{tr}} + \|\mathbf{R}_1 + \mathbf{R}_2\|_2)(\rho + \|\sum_{k=1}^2 (\mathbf{R}_k - \Phi_k)\|_2)} \end{aligned} \quad (38)$$

by applying Lemma 4 in Appendix A twice. The denominator in (38) is bounded, due to Assumption 1, and independent of λ . The numerator equals $\frac{1}{M} \|\mathbf{R}_1 - \lambda \mathbf{R}_2\|_F^2$. Hence, if (15) holds, then it follows from (38) that Assumption 2 also holds.

¹Observe that under Assumption 1 the matrices $\mathbf{Q}^{-1} \mathbf{R}_i \mathbf{Z}^{-1} \mathbf{R}_k$ have uniformly bounded spectral norm, which can be proved by using Lemma 3 in Appendix A.

APPENDIX D

We want to show that Assumption 2 implies that

$$\liminf_M \left(\beta_{11} - \frac{\beta_{12}^2}{\beta_{22}} \right) > 0. \quad (39)$$

Following the same line of reasoning as in [15, p. 24], Assumption 2 ensures (possibly over a converging subsequence, which exists because all quantities in (14) are bounded)

$$\lim_M \beta_{11} + \lambda^2 \lim_M \beta_{22} - 2\lambda \lim_M \beta_{12} > 0. \quad (40)$$

The left-hand side of (40) is a quadratic polynomial in λ and thus (40) is satisfied if and only if

$$\lim_M \beta_{11} + \lambda^2 \lim_M \beta_{22} - 2\lambda \lim_M \beta_{12} = 0 \quad (41)$$

has no real solutions. This requires the discriminant $\lim_M \beta_{12}^2 - \lim_M \beta_{11} \lim_M \beta_{22}$ to be negative, i.e., $\lim_M \beta_{12}^2 - \lim_M \beta_{11} \lim_M \beta_{22} < 0$, which together with $\lim_M \beta_{22} > 0$ leads to the desired result in (39).

REFERENCES

- [1] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, 2010.
- [2] E. G. Larsson, F. Tufvesson, O. Edfors, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Commun. Magazine*, vol. 52, no. 2, pp. 186–195, 2014.
- [3] J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. K. Soong, and J. C. Zhang, "What will 5G be?" *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1065–1082, 2014.
- [4] J. Hoydis, S. Ten Brink, and M. Debbah, "Massive MIMO in the UL/DL of cellular networks: How many antennas do we need?" *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 160–171, 2013.
- [5] H. Ngo, M. Matthaiou, and E. G. Larsson, "Performance analysis of large scale MU-MIMO with optimal linear receivers," in *Proc. IEEE Swe-CTW*, 2012, pp. 59–64.
- [6] X. Li, E. Björnson, E. G. Larsson, S. Zhou, and J. Wang, "Massive MIMO with multi-cell MMSE processing: Exploiting all pilots for interference suppression," *CoRR*, vol. abs/1505.03682, 2015. [Online]. Available: <http://arxiv.org/abs/1505.03682>
- [7] H. Yin, D. Gesbert, M. Filippou, and Y. Liu, "A coordinated approach to channel estimation in large-scale multiple-antenna systems," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 264–273, 2013.
- [8] A. Adhikary, J. Nam, J.-Y. Ahn, and G. Caire, "Joint spatial division and multiplexing—the large-scale array regime," *IEEE Trans. Inf. Theory*, vol. 59, no. 10, pp. 6441–6463, 2013.
- [9] X. Gao, O. Edfors, F. Rusek, and F. Tufvesson, "Massive MIMO performance evaluation based on measured propagation data," *IEEE Trans. Wireless Commun.*, vol. 14, no. 7, pp. 3899–3911, 2015.
- [10] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1993.
- [11] N. Shariati, E. Björnson, M. Bengtsson, and M. Debbah, "Low-complexity polynomial channel estimation in large-scale MIMO with arbitrary statistics," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 815–830, 2014.
- [12] X. Gao, O. Edfors, F. Tufvesson, and E. G. Larsson, "Massive MIMO in real propagation environments: Do all antennas contribute equally?" *IEEE Trans. Commun.*, vol. 63, no. 11, pp. 3917–3928, 2015.
- [13] R. Couillet and M. Debbah, *Random matrix methods for wireless communications*. Cambridge University Press, 2011.
- [14] A. W. Marshall, I. Olkin, and B. C. Arnold, *Inequalities: theory of majorization and its applications*, ser. Springer series in statistics. New York: Springer, 2011. [Online]. Available: <http://opac.inria.fr/record=b1132704>
- [15] F. Benaych-Georges and R. Couillet, "Spectral analysis of the Gram matrix of mixture models," *ESAIM: PS*, vol. 20, pp. 217–237, 2016.